

National Research University Higher School of Economics

As a manuscript

Zhukova Svetlana Yurievna

**Russian discourse formulae from a diachronic perspective (based on the data
from 18th -to 21st cent.)**

Dissertation Summary

for the purpose of obtaining

academic degree Doctor of Philosophy in Philology and Linguistics

Academic Supervisor

Doctor of Sciences in Language Theory, professor

Rakhilina Ekaterina Vladimirovna

Moscow, 2021

GENERAL CHARACTERISTICS OF THE THESIS

The new Construction Grammar theory that emerged in the last decades of the 20th century [Fillmore, 1988, 1989; Fillmore, Kay, 1992] changed many established views of modern science and posed a number of important large-scale tasks for linguists. Despite the fact that nowadays Construction Grammar includes plenty of concepts [see Hoffmann, Trousdale, 2013], there are some general principles that unite them. The most important of these is the acknowledgement of the absence of any clear boundaries between grammar and vocabulary. At some point along this border, there are constructions. Being a part of grammar, they are nevertheless subject to lexicographic description.

The task of compiling a list of constructions for each language, which was discussed already by Charles Fillmore [Fillmore, 2008], is one of the most urgent and laborious. Currently, constructions databases only exist for few languages (Swedish, German, Japanese etc.). An electronic database of constructions of Russian language, The Russian Constructicon, is being created as a part of a joint project of the Arctic University (Tromsø, Norway) and Higher School of Economics (Moscow, Russia). Describing constructions implies their formal as well as semantic properties being reflected, which makes it possible to automatically search within the database [Janda et al., forthc]. Another project dedicated to description of the Russian phraseological units is “Pragmacticon”, which is historically and ideologically related to “The Russian Constructicon”. It is also a database, but it aggregates and classifies non-one-word expressions of another type, which are called discursive formulas. These include expressions like *Vse jasno! Kak skažeš'! Tol'ko tak!* (‘Everything is clear! As you say! Only this way!’), which constitute the subject of the study. Historically, they were first detected when compiling the database of The Russian Constructicon. They were treated as special units that do not correspond to the format of the original database, but are still of indisputable interest both from a theoretical point of view and for teaching Russian as a foreign language.

By discursive formulas (DFs) we mean constructions that, unlike traditional ones, do not contain variables: they are fixed sequences of elements reproduced as a whole as a response to a verbal stimulus. And yet, these formulas have their own slots — they turn out to be moved outside, namely, into the previous and sometimes the subsequent segment. The term discursive formulae [perhaps use italics for term] itself goes back to the term speech formulae [perhaps use italics for term], introduced by Charles Fillmore in [Fillmore 1984]. The stability of the form and its position being fixed make it possible to compile a relatively complete list of DFs for the Russian language.

With regard to discursive formulas, the task of describing them from a diachronic point of view is complicated by the fact that, first of all, these units are used primarily in oral discourse. Thus, deep historical research on the material of the Russian language is impossible for them due to the genre limitations of the Old Russian texts. A natural limitation is the well-documented period of the 19th-21st century (with the involvement of data from 18th century), where spoken language is somehow conveyed in written texts — stories, plays, reports, etc.

The purpose of the work is to identify the dynamics of development and to outline the key patterns of change in discursive formulas of the Russian language during the 19th-21st centuries.

According to the goal set, the following challenges were met:

- 1) determining the theoretical boundaries of the concept of a discursive formula when comparing different approaches to similar linguistic phenomena;
- 2) compiling a list of discursive formulas sufficient for microdiachronic analysis;
- 3) analysis of the compiled list of discursive formulas from the historical dynamics point of view;
- 4) outlining the main pathways of change in the considered time period.

The novelty of the study lies in the fact that instead of using the traditional inductive approach and pointed observations of the history of particular units I consider

a representative list of DFs of the Russian language from the point of view of general trends in microdiachronic change in constructions of such kind. At the same time, studies on particular DFs serve as an illustration of the identified processes.

The history of the specific DFs requires the use of methods of corpus semantic and discourse analysis. At the same time, a comprehensive study of DFs is impossible without the use of natural language processing and the use of statistical **methodologies**¹.

The theoretical significance of the research is determined by the fact that the focus of its attention is on the material that has partly fallen into the peripheral zones of the grammar of the Russian language (cf., for example, word-sentences in [Shvedova 1980]), or (also partly) in the zone of attention of pointed lexicographic / lexicological observations (cf. [Sharonov 1996], [Melikyan 1999], [Baranov, Dobrovolsky 2000]). The work proposes, firstly, the definition of DF as a special observable class of constructions (which will further help to outline the system of such units in typological perspective), and secondly, its study from a dynamic point of view that is non-trivial for all these works. As a matter of fact, this means an opportunity to build DF issues into modern theories of constructionalization and pragmaticization [Lehmann 2002, 2004; Hopper, Traugott 2003; Bybee 2002, 2003, etc.].

The practical significance of the work is due to the fact that DFs are frequent in speech, but so far there is no representative list of them for the Russian language. At the same time, such a list can be used in teaching Russian as a foreign language, as well as in creating vocabulary resources of the Russian language and grammar of language change. The results of the work can be applied to improve the automatic communication of chat bots.

The following statements are presented for the defence:

¹ In Russian studies, A. Mustajoki was a pioneer in the use of such techniques, but applied to grammatical phenomena, cf. primarily [Mustajoki, Hannes Heino 1991].

1. DFs is a special class of constructions (from the point of view of Construction Grammar theory) with a set of specific properties that distinguish them from standard constructions.

2. In the Russian language (possibly, unlike some others) they constitute a very big class (hundreds of units), the study and classification of which is important not only for Russian grammar, but also for typology and theory of language.

3. DFs are the result of constructionalization and are a good example of pragmaticization. Different discursive formulas embody different stages of this process.

4. DFs are dynamic in nature. Significant shifts in their structure and semantics are observable already at an interval of 200 years.

5. Linguistic analysis with the use of statistical methods makes it possible to identify both productive patterns and general trends in changes in the structure of the list.

Approbation of the results of the work. The main results of the study were presented at the Constructicon Symposium HSE (March 17-21, 2017, Moscow), the Russian Verb conference (November 15-17, 2017, St. Petersburg), Kolmogorov seminar (April 25, 2018 Moscow), the international conference “Constructional semantics: Cognitive, functional and typological approaches” (August 24-25, 2018, Helsinki, Finland), the conference “Russian in the multilingual world” (April 10-12, 2019, Moscow), an international conference dedicated to the 110th anniversary of the birth of V. G. Admoni (October 8-10, 2019, Moscow). Based on the results of the study, six articles were published, four of which are included in journals from the list of the Higher Attestation Commission (VAK) and the list of NRU HSE.

The structure of the thesis. The thesis consists of an introduction, two chapters, a conclusion, a bibliography, and an appendix.

Chapter 1 includes 5 paragraphs. It is devoted to the theoretical problems of identifying discursive formulas as a special kind of constructions within the framework of Construction Grammar theory (§1, §2), the analysis of the description of formula

sequences in anglophone academic tradition (§3), and the analysis of the principles of forming registers of units similar to DF in the Russian tradition, taking the specifics of their description into account (§4). One of the aspects of such a description in the existing practice of studying lexical units and grammatical constructions is capturing the microdiachronic change (§5).

Chapter 2 is a systematic study of the list of Russian discursive formulas compiled in the course of the work based on the data of the National corpus of Russian language using statistical methods. It reveals a method for forming a list of discursive formulas of the Russian language (§1), which is further analyzed in terms of volume and structure (§2). The analysis of this list in the light of the general diachronic processes occurring with DF, using the techniques described §3 is presented in §4. In addition to analyzing the list as a whole, the paper investigates outdated and growing formulas with nominal (§5) and verbal (§6) anchor components, and §7 demonstrates the special role of the particles in the formation and dynamics of DFs. The conclusion summarizes the results of the work. The appendix presents a variant of the list consisting of 1249 units.

MAIN CONTENT OF THE DISSERTATION

§1 of Chapter 1 provides a brief overview of the terminological tools used in the work. The study follows the terminology adopted in Construction Grammar [Fillmore et al. 1988]. The term ‘construction’ is used to refer to a linguistic expression that has form and meaning, which cannot be derived from the meaning or form of its constituents [Goldberg, 1995. C. 4].

Constructions include constant elements and slots (positions) that can be filled with different variables depending on the relevant semantic and grammatical constraints. Constructions are considered at different levels, including discursive and pragmatic [cf. Croft and Cruse 2004]. §2 substantiates the allocation of discursive formulas (DF) as non-one-word non-compositional units like: *Èto ešče čto! Vot ono kak! A to!* (‘What’s this! That’s how it is! You bet!’) As a special peripheral class of

constructions, DFs constitute complete cues belonging to a single speaker. Their main feature is that, unlike conventional constructions, they do not contain variables within the formula itself. DFs are reactions to a verbal stimulus of a certain type, and the preceding speech act acts as a slot for them. At the same time, the original lexical meaning of the DF turns out to be bleached, and the pragmatic meaning ends up as the main one for formulas of this type.

The following sections discuss the place of these units in different linguistic traditions. §3 discusses the anglophone tradition of identifying formulaic sequences. Formulaic sequences refer to continuous or discontinuous sequences of words or other elements of the language, which seem to be prepared in advance: they are stored in memory as a whole and are extracted directly at the moment of speaking [Wray 2002]. Against this background, DFs are seen as a special type of formulaic sequences: they are non-compositional, non-one-word, easily reproducible, represent a complete cue and have minimal structural variations, they are a response of a certain type — and, apparently, are also stored entirely and extracted at the moment of speaking.

§4 is devoted to the Russian-speaking tradition of describing formula units and approaches to forming a list of them for Russian language. It is shown that the units similar to DF fell into the focus of attention of researchers very early — back in the middle of the 20th century, cf. firstly [Shvedova 1960]. However, the tradition of describing these units has not been formed — for instance, in Russian Grammar-80 only one section is reserved for them, where they, in fact, are put into a category of peripheral syntactic phenomena. Currently, it is possible to distinguish various approaches and the corresponding terms, which in one way or another serve this zone of the Russian language. The main ones include speech formulas (section 4.2), communicemes (section 4.3) and communicatives (section 4.4).

The term “speech formulas” was introduced and widely used by theorists and practitioners of Russian phraseology — A. N. Baranov and D. O. Dobrovolsky. These units are defined as a subtype of idioms, “which in form are independent sentences, and in function they represent a reaction to previous actions (including statements) of the speaker themselves or another participant in the situation [Baranov, Dobrovolsky

2000]. The class of speech formulas is similar to DF by the fact that the units corresponding to it can also serve as responses in a dialogue. However, unlike DF, speech formulas include reactions not only to a specific fragment of discourse (a cue of an interlocutor), but also to actions. Another difference between speech and DF is the fact that the latter appear exclusively as complete cues, while speech formulas can be built into the structure of a sentence. Differences in the use of units of both groups are associated, among other things, with their semantic features: if DFs are variations of positive or negative answers, and are also evaluations, then speech formulas can also be used to express accompanying comments (*ne laptem šči hlebaem* ‘we don’t slurp soup with a shoe’). In general, DFs can be regarded as a special subtype of speech formulas, which are to some extent included in the Dictionary-Thesaurus of Modern Russian Idioms [Baranov, Dobrovolsky 2007]. Indeed, several dozen DFs can be found among the speech formulas there — but for obvious reasons they are not outlined as a separate type.

The notion of “communicema” was introduced by V. Yu. Melikyan in his “Dictionary of emotionally expressive turns of living speech” [Melikyan 2001]. The dictionary is compiled on the basis of explanatory and phraseological words of the literary and abusive language with the involvement of data from fiction and journalistic literature of the 19th-20th centuries. The list of those included in the dictionary is very heterogeneous; as shown in 4.3, only about a third of 750 can be considered the actual DFs, so it is far from complete in terms of DF.

Communicatives (Section 4.4) are a class of special units which is the closest to DF. The term “communicative” was introduced into scientific use a long time ago — it was used by V. F. Kipriyanov in his work “Phraseologisms — communicatives in the modern Russian language” [Kipriyanov 1975], and later consolidated in the works of Sharonov [Sharonov 1996 et al].

There are many similarities in the definition of DFs and communicatives. Communicatives, like DFs, are separate cues that act as a reaction to what the interlocutor has said. I. A. Sharonov considers communicatives to be the responses which depend on the previous ones in such a minimal unit of a dialogue as “an adjacent

pair of phrases of both speakers”, in which the first is a stimulus and the second is a response [Sharonov 2014]. It is this that conceptually brings them closer to DFs, because discursive formulas are also response cues to a verbal stimulus. The diachronic aspect of functioning of DFs and communicatives is of mutual interest. However, the systematic description of communicatives has not been completed yet, there is still no full list of them, and the examples presented in individual works are very heterogeneous. Apparently, the class of communicatives includes, among others, units that are not similar to DF — for example, interjections, so the restrictions on the structure and semantics of the communicatives are generally weaker than on DF.

Having determined the place of DFs in area of formulaic idioms, we proceed to discussing the niche they occupy in diachronic studies. Within the framework of Construction Grammar, the focus of attention always ultimately turns out to be the history of a separate construction on language material fairly deeply rooted in history, taking statistics into account [Hilpert 2013]. For the Russian language, there is no such possibility (except for superfrequent units) — so far, we can only talk about microdiachronic studies (§5). Microdiachrony refers to the “shallow” historical layer of the Russian language: the change that took place in it over the past 200-300 years. This change is well-documented in the RNC, as it is researched using corpora and in relation to particular lexemes cf. [Daniel, Dobrushina 2016], and for large data sets [Kuzmenko, Kutuzov 2017], and for studying the texts of 19th century that are distant from us in time and in relation to the linguistic norm [Rakhilina et al. 2016, Rakhilina 2017]. All these publications serve as a methodological basis for the conducted microdiachronic research of DF as lexical and grammatical units.

Narrowing data down to microdiachrony is not fatal: as shown in Section 5.4, the study of DFs, even on the material of a relatively short time interval of 150-200 years, allows tracing significant change within the framework of such processes as grammaticalization and pragmaticization. Pragmaticization is a process in which a linguistic unit, in a certain context, changes its propositional meaning to predominantly metacommunicative and discursive [Frank Job 2006: 397] and takes the form of independent statements. In the Russian tradition, the term “pragmaticization” in

context of studying the units of oral discourse is used primarily in the works of the group of N. V. Bogdanova-Beglaryan within the project “One Speech Day” of the Spoken Corpus of the Russian language [Bogdanova-Beglaryan 2021 et al]. It is as a result of this process that lexemes or whole constructions become pragmatic markers. However, the class of pragmatic markers per se does not directly correlate with DF: these markers perform different functions and are not used as separate cues. The similarity here is that the term “pragmaticization” is also applicable to DFs, since these are also units with a special, pragmatic meaning — emphatic consent, approval, denial, etc.

Chapter 2 is entirely devoted to the study of linguistic material and analysis of the dynamics of microdiachronic change in DF with the use of the National Corpus of the Russian Language. An important achievement of the work is introducing the list of DFs for the Russian language, a version of which was used as the basis for the Pragmaticon Database (pragmaticon.ruscorpora.ru). Therefore, the first sections of the second chapter are devoted to the process of creating such a list, its discussion and diachronic work with it.

§1 of Chapter 2 describes the process of creating an automatic module system for extracting the discursive formulas of the Russian language. The linguistic material for this module was dramatic texts, which to the greatest extent reflect dialogic oral speech. The module made it possible to automatically select DF cues based on their formal features. In the course of the work, a corpus of manually-annotated dramatic texts was developed, which was then used to train the system. After that, using the program, 420 plays of a period from 19th to 21st century taken from two corpuses of dramatic texts were processed [Lubimovka 2018, Russian Drama Corpus 2018]. As a result, after additional manual processing, a list of more than 3000 units was obtained. Clearly, such a list was too long and heterogeneous. §2 discusses classes of units that were systematically excluded from this list. In particular, these are *routines* — stable units that represent a response not to a verbal stimulus, but to some physical element of a communicative situation (cf.: *čto že èto delaetsja* ‘what’s happening’; *da-da, vojдите* ‘yes, yes, enter’). Moreover, there are *discursives* — units that do not convey a

yes / no answer and do not express an evaluation, but instead participate in regulating the dialogue. They can be speech stimuli aimed at continuing the story (cf. *čto dal'she?* 'what's next?') or at stopping it (cf.: *nu hvatit uže* 'well, enough already'), to mark returning to the old topic (cf.: *vernemsja k našim baranam* 'let's get back to the point (lit. to our sheep)') etc. Note that such units themselves are of indisputable linguistic interest, so that this linguistic material can serve as a basis for further research.

After a number of manipulations, including combining structural variants of one DF, a list of about 1000 units was obtained. In order to ensure the possibility of using statistical methods, the frequency of DFs in the texts of RNC was determined for each time interval of 10 years. It is described in more detail in Section 3.1.

As a result, a table was composed, the initial analysis of which made it possible to confirm the idea of these units being highly dynamic in nature. They can be somewhat conventionally divided into three groups:

- 1) outdated and close to becoming outdated (cf.: *da čto takoe* 'what is it', *kak ty smeeš'* 'how dare you', *sdelaj milost'* 'do me a favor'),
- 2) growing — appearing from the middle of the 20th — beginning of the 21st century (cf.: *bez problem* 'no problem', *vse v porjadke* 'everything is alright', *tak ne byvaet* 'it is not possible'),
- 3) conditionally stable, which are uttered with one or another measure of constancy throughout almost entire analyzed period (cf. *čto takoe* 'what is it', *čto èto* 'what is it', *kak že* 'how is it').

A number of methods of statistical analysis were applied to the data obtained, including the calculation of the absolute and relative growth rate of each unit. A positive growth rate indicates a tendency towards an increase in the frequency of DF use, while a negative growth rate may indicate a tendency for DF to go out of use. The relative growth shows the dynamics of change in the composition of the list of each DF in relation to the other.

Clearly, obsolescence can occur due to the withdrawal from use of one or another word in a DF (or a meaning of a word) and replacing it with another. These processes were illustrated in the thesis using the examples of formulas *počem ja znaju* ‘how do I know’ (with *počem* being obsolete in the corresponding meaning) and *otkuda ja znaju* ‘how do I know’.

In addition to the obsolescence of individual words that constitute a DF, the work considers the process of absorption of the formula by its more frequent synonymous version, as it happened with the DF *kak tak* ‘how come’ which was replaced by the DF *kak èto* ‘how come’ (Section 4.1.2). It is usually indicated in dictionaries that *kak tak* and *kak èto* should be considered variants of the same linguistic unit. It could be assumed that the dependency of their use on each other is direct: the growth or decline of one unit entails a similar behavior of the other. However, it was not statistically confirmed — the relationship of the use of these DFs is rather inverse: initially, *kak tak* is the one that dominates, then *kak tak* shows a decline in frequency, and after that *kak èto*, on the contrary, shows growth. It turns out that variants of one formula can replace each other over time.

This statement — the hypothesis of “minimal structural variation” — was decided to be verified by statistical estimates on extended data (Section 4.1.3). The study design was as follows. For the study, DFs with variability were selected, such as *a kak inače* ‘how else’, *a kak že inače* ‘how else’; *tak čto ž* ‘so?’, *nu tak čto* ‘so?’. Then, for each of the DF variants, the values of entries in the corpus for individual decades were obtained, and then the correlation coefficient was calculated for each pair of variants. It was expected that competing forms would receive a high negative value of this coefficient. This would mean that at the moment when the frequency of one DF variant grows, the frequency of the opposed variant should fall and vice versa. The negative value of the coefficient itself should have confirmed the systemic relationships of the DF variants in time.

However, on the basis of the available small data, the hypothesis was only partially confirmed. It turned out that during the entire period of more than 200 years, the DF variants tend to exist simultaneously, apparently complementing each other. At

the same time, it can be seen that the DF variants replace each other in short-term periods of their coexistence.

The next part, Section 4.2, is devoted to the general processes that form new DFs. As shown by the example of the DF *ne fakt* ‘not a fact’ the formation of new DFs can occur on the basis of full-fledged constructions — in particular, through reducing valency, which is expressed by a content clause. Cf.: *ne fakt, čto P* [construction with a *P* variable] – $\rightarrow P$. – *Ne fakt* (isolated DF) (Section 4.2.1). Another important process influencing the appearance of DFs is calquing — its features are considered using the example of the formula *bez problem* ‘no problem’ (Section 4.2.2).

Thus, §4 reflects general processes of microdiachronic change in DF. The next part of the work is devoted to how the support word of the anchor component of DF affects this change — a noun (§5), a verb (§6) or particles (§7), which, as demonstrated in the work, can also be considered a significant element of the anchor component of DF.

Nouns in the anchor component of the DF almost always have the potential for variation — and by this they lay the foundation for the change in DF over time. Among the anchor nouns, words related to religious themes stand out as especially frequent ones (cf.: *Bog s tobj!* ‘God be with you’] or: *Gospod' ego znaet!* ‘who knows (lit. God knows him)’. It turns out that DFs with such lexemes show negative growth. However, the models themselves, according to which they are formed, are extremely productive, since other lexemes can be embedded in them, including rude vernacular ones such as *fig, hren* as well as obscene vocabulary.

As the study shows, DFs with nouns are predominantly evaluative formulas, although there are also agreement / confirmation and disagreement formulas among them that are particularly interesting to us.

As an example of an agreement formula, the DFs *delo / volja tvoe / vaše* ‘it's your call’ are considered. They are synonymous, but their diachronic behavior is different. The formula with the lexeme *volja* ‘will’ shows a decline, while the frequency of DF with the lexeme *дело*, on the contrary, is increasing.

An example of conformation is the X-y formula *ponjatno* ‘understood’, and also the X-y formula *jasno* ‘(it’s) clear’, which is close to it in its structure and semantics. They coexist and both allow embedding as a situational “comprehensibility reference point” for evaluative lexemes such as *durak* ‘fool’. The formal difference between them is that the X-y *ponjatno*, to a much greater extent than the X-y formula *jasno*, tends to include the names of animals, importantly, ones that are rare and lack any connotation, negative or positive, cf.: *ež* ‘hedgehog’ or *koza* ‘goat’. Apparently, the fact that the nomination for the “reference point” for the subject of perception and understanding the information has been chosen in such a non-trivial way, enhances the shade of evaluativeness and aggression in the formulaic answer X-y *ponjatno*. At the same time, the X-y constructions *ponjatno / jasno, [P]*, comparable to DFs, develop: the pronouns *ljuboj / každyj* ‘any / every’ begin to be embedded into them. Quantification makes the statement even more intense. Then, the quantifier pronoun can be substantivized. New variants of constructions become the basis for new DFs. Thus, the case of these formulas show the process of emergence of a DF at the stage when the original syntactic construction itself is ΠT the process of formation.

An example of a disagreement formula (with a clear evaluative component) is the DF *(ty) s uma sošel / spjatil* ‘(you) have gone crazy’. It goes back to a rhetorical question (cf.: *Ty s uma sošel?* ‘Have you gone crazy?’), from which the disagreement meaning with a clear negative evaluation by the speaker is derived. This formula coexists with another evaluative formula, which is very similar to it in structure: *s uma sojti (možno)* ‘it is crazy (lit. you can go crazy)’. In both cases, a reduction of the structure is observed — in the formula *(ty) s uma sošel / spjatil*, pro drop takes place, and the DF *s uma sojti (možno)* turns out to be a free infinitive [Paducheva 2019], which has “lost” its modal verb. It is shown that these formulas, which are very similar at first glance, have not only different nature and different structures, but also different semantics (respectively, disagreement, with a negative evaluation by the speaker and context-based evaluation (positive or negative)).

§6 discusses the dynamics within a list of formulas with verb anchor component. Verb formulas make up about 40% of the analyzed list. The most frequent ones turned

out to be the verbs of mental activity and speech: *znat'* 'know' (20%), *govorit'* 'talk' (11%), *skazat'* 'say' (7%), *dumat'* 'think' (6%); existential verb *byt'* 'be' (10%); modal verb *hotet'* 'want' (5%).

As an example of canonical DFs with the elements of fixed form, this section discusses formulas *kak hočeš'* 'as you wish (lit. want)' and *kak znaeš'* 'as you wish (lit. know)'. Both of them can have both the usual two-part structure or a more complex three-part structure. Both two-part constructions express consent, which is nonetheless different in each case. The formula *kak znaeš'* expresses forced consent, which may be accompanied by resentment or internal rejection of the information provided, since the interlocutor made a decision regardless of the speaker's opinion. The formula *kak hočeš'* conveys a voluntary consent, when the speaker independently delegates the responsibility for the choice being made to their interlocutor.

The three-part construction has the following structure: the first cue belongs to the speaker, the second cue belongs to their interlocutor and represents a refusal, and the third one belongs to the speaker again, as he reacts to their interlocutor's refusal. At first glance, they seem to be synonymous, but in fact they have important differences in use. The starting point for the DF *kak hočeš'* 'as you wish (lit. want)' are speech acts of offering to help, suggestions, and advice. At the same time, for the DF *kak znaeš'* 'as you wish (lit. know)' the range of opening speech acts would be somewhat different: a request, a request for permission, or an offer to help. As a stimulus, the speaker can receive not only a refusal, but also a prohibition. It turns out that in this case DFs *kak hočeš'* *kak znaeš'* mark different speech acts.

It is important to note that the key verbs in both growing and obsoleting formulas are still *znat'* 'know', *govorit'* 'talk', *skazat'* 'say'. This suggests that the class of verb formulas is fairly stable. Minor changes have occurred since the 1980s, when more coarse vocabulary such as *gnat'* 'tell' appear.

The material considered in the work showed certain trends, so, with some conventions made, factors influencing the dynamics of change in verbal DFs can be discussed. Thus, the change in order of words inside a DF turns out to be significant. For example, the formula *a čto ja govorju / govoril* 'and what do / did I say' has a

tendency to fall out of common use, while the frequency of the use of the variant *a ja čto govorju / govoril* ‘and what did I say’ with a pronounced phrasal stress is growing. This word order is labeled pragmatically, as it only reinforces the meaning of reproach.

The dynamics of change can also be influenced by the replacement of the polite form of the second person pronoun *vy* with the pronoun *ty*: (*čto vy / ty takoe govorite / govoriš* ‘what are you saying’). Of course, the choice depends on the social status of the participant in the situation, however, you can notice a tendency to consolidate less polite form of the second person singular pronoun.

But the fact that verbal DFs are characterized by fixation of a verb in a certain form with possible variability in number, there are some formulae in which variability in other parameter is observed — for instance, the personal form of the verb can be replaced by an infinitive, as in case of the formula *čto tut podelaesh’ / podelat’* ‘what can you do about it’. There is also a noticeable tendency to replace the form of the present tense verb with the form of the past tense (*a ty čto hočesh’ / hotel* ‘and what do / did you want’, *a ty kak dumaeš’ / dumal* ‘and what do / did you think’ etc.)

At first glance, it seems that, unlike nouns and verbs, particles cannot be considered with full-valued anchor components. Indeed, the original lexical meaning of the particles is erased and to a large extent it is the particles that provide the variability of the formula, while full-valued verbs and nouns tend to serve as an anchor constant for a DF. However, as it is shown in the work, particles often act as constituent components of the anchor part of a formula, determining its semantics and the direction for diachronic change — therefore, separate section (§7) is devoted to them in the series of the previous two.

This section discusses the most common particles in the DFs: *a, da, nu*. Particles generally constitute a rather small class of units that are interesting from the point of view of their functionality, cf. [Baranov 1987, Dobrushina 1994, 1995] and many others, translation strategies [Dobrovolsky, Levontina 2017], variability [Valova 2016] etc. The objective of the current study is different: it suggests the analysis of microdiachronic change in the use of these particles within the DFs.

It can be observed that formulas containing particles are in general quite stable in historical perspective: throughout the entire period under consideration, their relationship with DF which do not contain particles, does not change. This is true for individual particles *tak*, *da*, *že*, *a*. The particle *nu* ‘well’ follows a rather special behavior. Judging by statistical calculations, the use of formulas with a particle *nu* turns out to be more preferable over time (Section 7.3). This makes it stand out from the rest of the particles. The peculiarities of the formation of DFs built on the basis of the *nu* particle are discussed in Section 7.5. The case of the particle *nu eště* demonstrates the process of emergence of a formula in its first stages of pragmaticization, unlike the *nu-nu* and *nu i nu*, which are used as full-fledged DFs already in the texts from the first half of the 19th century. It is also interesting that the particle *nu* replaces the particle *a* in historical perspective — in such a way that the initial meaning of the formula *a čto* ‘so?’, characteristic of the texts of the 19th century, is currently realized through a formula with the initial *nu*.

The role of particle in a DF can be different — this is discussed in Section 7.2. As already mentioned, the particles themselves can form a DF (*to-to*, *nu-nu*). It is not uncommon that creating a DF happens precisely by adding a particle to it (cf. *i vse / nu i vse* ‘that’s it / well that’s it’, *tak čto / tak čto že* ‘so what’, *znaeš' / nu znaeš* ‘you know / well you know’ etc.). Moreover, the attachment of a particle can change the semantics of a DF, including to virtually the opposite. It happens, for example, to the DF *kak že*, which expresses disagreement, but with the addition of the particle *a* begins to express a confident confirmation. Particles can also affect shifts in the use of a formula — a good example is the history of the DF *nu / da čto vy govorite* ‘well / is that right (lit. what are you saying)’. Initially synonymous, these DFs are not different in their meaning, and each of them turn out to be assigned to a certain particle. All this suggests that the distribution of formulas by particles is not arbitrary and the replacement of a particle, which, as it may seem, generates a free variant of a DF, in fact most likely changes the properties of this formula significantly.

Thus, when examining DFs, it is important to take into account not only their functions, but also their structure. In Russian, it has been shown to be determined by

the units of three functional-semantic classes: nouns, verbs, and particles. Further research will show its relevance in a typological perspective.

The Conclusion summarizes the results of the study.

The work substantiates the existence of a special class of markers with pragmatic meaning — discursive formulas (DF) — against the background of a detailed analysis of various kinds of discourse units that are traditionally defined in Russian studies and general linguistic tradition. These are non-one-word responses in a dialogue, expressing the meaning of *da / net* or a value judgement. It is shown that the importance of the research on this kind of units is still underestimated, although it is of both theoretical and practical interest, especially in view of them being quite numerous.

The proposed approach helps to clarify the nature of dfs and to look at them in the light of such processes as construalization and pragmaticization. For rapidly “bleached” discursive units, these processes can be observed over a relatively short period of time, well documented in RNC, starting from the 18th century. The pace of the processes is different: in general, it can be stated that DFs that appeared in the early or middle of the 19th century to the beginning of the 20th century function as pragmatic units. Nevertheless, there are those which already in the early texts show a high degree of pragmaticization, as well as those which are currently going through this process.

Therefore, this work opens up a new area of research, offering the scientific and pedagogical community not only a detailed list of Russian DFs (which later formed the basis of the public resource “Pragmaticon” pragmaticon.ruscorpora.ru), but also a general idea of the dynamic processes that apply to them. This idea is formed, on one hand, on the basis of statistical studies, and on the other, from the analysis of semantics, history of use and variability of specific formulae, such as: *ne govori / ne skaži, kak hočeš' / kak znaeš', X-y ponjatno, tak čto že, a vy govorite* etc.

The main content of the work is reflected in the following **publications**:

- 1) Puzhaeva S. Yu., Gerasimenko E. A., Zakharova E. S., Rakhilina E. V. Avtomatičeskoe izvlečenie diskursivnyh formul iz tekstov na russkom jazyke [Automatic extraction of discourse formulas from texts in Russian.] // Bulletin of Novosibirsk State University: a series of linguistics and intercultural communication. T.16, No. 2, 2018.
- 2) Zhukova S. Yu., Orekhov B. V., Rakhilina E. V.. Diskursivnye formuly russkogo jazyka: diahroničeskij podhod [Discursive formulas of the Russian language: a diachronic approach] // Proceedings of the Institute of the Russian language named after VV Vinogradov. – 2019. – T. 21. – C. 142-164.
- 3) Zhukova S. Yu. Diskursivnye formuly russkogo jazyka kak hočeš', kak znaeš' v diahroničeskom aspekte [Discursive formulas of the Russian language as you want, as you know in the diachronic aspect] //Acta Linguistica Petropolitana. Proceedings of the Institute for Linguistic Research. 2019.T. 3. №. XV.
- 4) Rakhilina E.V., Bychkova P.A., Zhukova S. Yu. Rečevye akty kak lingvističeskaja kategorija: diskursivnye formuly [Speech acts as a linguistic category: discursive formulas] //Topics in the study of language. 2021. №. 2. C. 7-27.
- 5) Defining discourse formulae: computational approach. Ekaterina Gerasimenko , Svetlana Puzhaeva , Elena Zakharova , and Ekaterina Rakhilina //EPiC Series in Language and Linguistics Volume 4. 2019. Pages 61–69.
- 6) Rakhilina E. V., Plungyan V. A., Puzhaeva S. Yu. “ Ob ètom rano govorit'...” [It's too early to talk about it ...] // D. V. Gerasimov, S. Yu. Dmitrenko, N. M. Zaika (ed.). Collection of articles dedicated to the 85th anniversary of V. S. Khrakovsky, JASK Publishing House, 2019. 5) "It's too early to talk about it ..." //, 2019.